

From Word Models to World Models: Translating from Natural Language to the Probabilistic Language of Thought

Year: 2023 | Citations: 125 | Authors: L. Wong, Gabriel Grand, Alexander K. Lew, Noah D. Goodman, Vikash K. Mansinghka

Abstract

How does language inform our downstream thinking? In particular, how do humans make meaning from language--and how can we leverage a theory of linguistic meaning to build machines that think in more human-like ways? In this paper, we propose rational meaning construction, a computational framework for language-informed thinking that combines neural language models with probabilistic models for rational inference. We frame linguistic meaning as a context-sensitive mapping from natural language into a probabilistic language of thought (PLoT)--a general-purpose symbolic substrate for generative world modeling. Our architecture integrates two computational tools that have not previously come together: we model thinking with probabilistic programs, an expressive representation for commonsense reasoning; and we model meaning construction with large language models (LLMs), which support broad-coverage translation from natural language utterances to code expressions in a probabilistic programming language. We illustrate our framework through examples covering four core domains from cognitive science: probabilistic reasoning, logical and relational reasoning, visual and physical reasoning, and social reasoning. In each, we show that LLMs can generate context-sensitive translations that capture pragmatically-appropriate linguistic meanings, while Bayesian inference with the generated programs supports coherent and robust commonsense reasoning. We extend our framework to integrate cognitively-motivated symbolic modules (physics simulators, graphics engines, and planning algorithms) to provide a unified commonsense thinking interface from language. Finally, we explore how language can drive the construction of world models themselves. We hope this work will provide a roadmap towards cognitive models and AI systems that synthesize the insights of both modern and classical computational perspectives.